J Orthop Sci (2014) 19:242–249
DOI 10.1007/s00776-013-0517-8

ORIGINAL ARTICLE

Measurement properties of the Dutch version of the Western Ontario Shoulder Instability Index (WOSI)

Suzanne H. Wiertsema · Pieter Bas de Witte · Marc B. Rietberg · Karin M. Hekman · Maaike Schothorst · Martijn P. Steultjens · Joost Dekker

Received: 27 September 2013 / Accepted: 2 December 2013 / Published online: 25 December 2013

Abstract
Background The Western Ontario Shoulder Instability Index (WOSI) is a disease-specific shoulder questionnaire to measure quality of life in patients with shoulder instability. The aim of the present study was to translate the WOSI into Dutch and assess its principal measurement properties.
Methods The WOSI was translated into Dutch according to guidelines in the literature. Fifty-two shoulder instability patients completed the questionnaire twice within 2 weeks. We assessed internal consistency (Cronbach's alpha), test–retest reliability [Intraclass Correlation Coefficient (ICC)], standard error of measurement (SEM), smallest detectable change (SDC) and reliable change index. The Bland–Altman analysis was applied to assess test–retest agreement and floor and ceiling effects were calculated.
Results Cronbach's alpha was 0.95 for the total WOSI score (range 0.88–0.95 for the 4 domains). ICC for the total WOSI score was 0.91 (range 0.79–0.90 for domains), SEM was 130.6 for the total WOSI score resulting in a SDC of 362.0, which is 17.3 % of the maximum obtainable score of 2100. Bland–Altman analysis showed no systematic differences or consistent bias between the two assessments. We observed no relevant floor and ceiling effects.
Conclusion The results of the present study suggest the Dutch version of the WOSI is a reliable tool for clinical assessment and scientific evaluation. It shows high values for Cronbach's alpha and ICC implying excellent internal consistency and good test–retest reliability.

S. H. Wiertsema (✉) · M. B. Rietberg · K. M. Hekman · M. Schothorst
Department of Rehabilitation Medicine, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands
e-mail: s.wiertsema@vumc.nl

P. B. de Witte
Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands

M. P. Steultjens
Institute for Applied Health Research and School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK

J. Dekker
Department of Rehabilitation Medicine and EMGO Institute, VU University Medical Center, Amsterdam, The Netherlands

Introduction

Shoulder instability is a common shoulder disorder, mainly affecting young individuals [1]. Shoulder instability can have a variety of origins, but the majority of patients have had a traumatic anterior glenohumeral dislocation, which has a lifetime risk between 1 and 2 % [2, 3]. Recurrent instability is the most frequent complication after a first acute traumatic luxation [4]. There are numerous studies on the diagnosis and treatment of shoulder instability, but there is a need for well-validated outcome measurements focussing on shoulder instability .

One of the most essential factors determining treatment outcome is how the patient perceives his own health status [5]. As a result, many self-reported questionnaires that measure health-related quality of life (HR-QOL) have been developed over the past decades. It is well-known from the literature that disease-specific instruments assessing HR-QOL are more accurate in measuring changes related to specific disorders than general instruments [6, 7]. For accurate patient assessment it is recommended to combine a general health outcome measurement, a regional outcome measurement and a disease-specific outcome measurement

[8]. For shoulder instability this means that it is essential to use at least a disease-specific evaluation tool that assesses, e.g., apprehension and confidence in the shoulder in addition to pain, strength, activities above shoulder level and range of motion, which are important items in most general shoulder scores [9].

As a consequence of the increasing use of HR-QOL questionnaires in a clinical setting and in research, growing interest in the measurement properties of these questionnaires has been observed.

The Western Ontario Shoulder Instability Index (WOSI) was introduced by Kirkley and colleagues in 1998 [2]. This questionnaire is a self-reported disease-specific outcome measurement to assess HR-QOL in patients with shoulder instability. In recent studies comparing validated self-reported shoulder instability scores, the WOSI was reported to have the best measurement properties [8, 9].

The WOSI is an increasingly applied outcome measurement in clinical shoulder instability studies. Over the past years, the WOSI has been translated and well validated for use in Sweden, Germany, Italy and Japan [10–14]. We found one follow-up study using a WOSI in the Dutch population; however, the translation process and measurement properties of the applied WOSI were not described [15]. There is a need for a Dutch translation of the WOSI, translated according to international guidelines and for which measurement properties are well investigated. In the present study we cross culturally adapt the WOSI for use in the Netherlands and determine its reliability in terms of internal consistency, test–retest reliability and measurement error, according to the COSMIN taxonomy (COnsensus-based Standards for the selection of health Measurements INstruments) [16].

Materials and methods

Western Ontario Shoulder Instability Index

The WOSI assesses HR-QOL in patients with shoulder instability [2]. It is a self-reported questionnaire consisting of 21 items in 4 domains: physical symptoms (10 items); sports, recreation and work (4 items); lifestyle (4 items) and emotions (3 items). Each item is scored on a 100 mm visual analogue scale (VAS). Total score ranges from 0 to 2100, with higher scores indicating a reduced HR-QOL. The WOSI score contains extensive written instruction for users, which includes a clarification of every single question.

Translation

Following approval of S. Griffin, one of the designers of the WOSI, translation of the questionnaire was performed according to guidelines in the literature [17]. The questionnaire was not translated literally, but a stepwise procedure was followed to achieve a conceptual translation. Steps include forward translation, reconciliation meeting, backward translation, comparison with source questionnaire, review by clinicians, debriefing and report. Forward translation from English to Dutch was done by 3 independent individuals; 1 physical therapist (KMH), 1 orthopaedic surgeon and 1 epidemiologist (MPS). It was back translated to English by a native speaker who is working as an occupational therapist in one of the participating hospitals. The back translated version was then reviewed by the three forward translators mentioned before, and compared with the original source. In a final consensus meeting the final Dutch version of the WOSI was agreed upon.

Patients

Patients were recruited from two university medical centers in The Netherlands: the VU University Medical Center (VUmc) in Amsterdam and the Leiden University Medical Center (LUMC). Both centers have specialized shoulder groups, in which orthopaedic surgeons and physical therapists work closely together in treating shoulder patients and performing shoulder-related research.

Eligible patients, diagnosed with shoulder instability, were identified from databases of patients who visited the orthopaedic outpatient clinic of VUmc or LUMC from 2009 to 2011. Inclusion criteria were: (1) older than 18 years, (2) current shoulder instability; traumatic, non traumatic, or post surgery and (3) shoulder pathology (e.g., dislocation, Bankart lesion) confirmed by radiological evaluation recently or in the past. Both operatively and non-operatively treated patients were included. In the case of surgery, the operation took place at least 6 months prior to inclusion, assuring rehabilitation was completed and a stable situation was achieved, which is essential for underlying reliability study. Patients with fractures, neurological disorders leading to shoulder symptoms, tumours, infections, cognitive impairments and patients with signs of cervical syndrome were excluded. All patients gave informed consent. The local medical ethics committees of VUmc and LUMC approved the present study. The minimal required number of patients to be included was 50, since it has been determined that this is an appropriate sample size to assess reliability parameters in health status questionnaires [18].

After screening the databases of the orthopaedic outpatient clinics of both VUmc and LUMC, 158 patients were considered eligible and were sent an information letter. After the 2 weeks reflection period a total of 34 patients could not be reached, 38 patients ultimately did not

Table 1 Participants characteristics ($N = 52$) and mean values of the total WOSI scores and the scores on the 4 domains

Variables	Mean ± SD or frequency (%)
Age	31.0 ± 10.1 (range 20–68)
Gender	
Female	19 (36.5 %)
Male	33 (63.5 %)
Type of shoulder instability	
Recurrent instability (traumatic)	22 (42.3 %)
Recurrent instability (non-traumatic)	10 (19.2 %)
Recurrent instability (post surgery)	20 (38.5 %)
Affected side	
Left	24 (46.2 %)
Right	27 (51.9 %)
Both	1 (1.9 %)
Time interval (days) between measurements	25 ± 17.3 (10–100)
WOSI score (min–max)	
WOSI total score	731.3 ± 435.4 (59–1841)
Physical symptoms	314.8 ± 199.4 (41–743)
Sports, recreation and work	157.7 ± 100.8 (3–399)
Lifestyle	133.6 ± 94.7 (6–391)
Emotions	127.4 ± 81.9 (4–300)

Mean WOSI scores are calculated over the outcomes of 2 measurements

met the inclusion criteria and were excluded for the following reasons: 22 no longer had complaints of instability, 7 had co-morbidity such as actual contusion or fractures, 3 had a limited ability to speak Dutch, 5 declined to participate and 1 was deceased. The remaining 86 patients received the WOSI questionnaire according to the study protocol. A total of 52 patients (33 men and 19 women with mean age of 30.9 years) completed the WOSI questionnaire twice and their data were included in the analyses (response rate 60.5 %).

In Table 1 characteristics of the study population, including type of shoulder instability are summarized. In addition the total WOSI score and the scores on the 4 domains are presented in Table 1.

Procedure

All eligible patients identified from the databases of the orthopaedic outpatient clinics of VUmc and LUMC received an information letter, and after a reflection period of 2 weeks, they were contacted by phone by one of the coordinating investigators (SHW, PBW). At that point, patients received further information, and inclusion and exclusion criteria were verified. Subjects, willing and eligible to participate, received the informed consent (IC) form and the first WOSI questionnaire by regular mail. The questionnaire and IC form were filled out at home and returned to the examiners, using pre-paid return envelopes. Patients were instructed to fill out the questionnaire without any help. Patients who did not respond within 2 weeks were contacted by phone by one of the coordinating investigators. Two weeks after initial response participants received the second WOSI with similar instructions as for the first WOSI. The time span of 2 weeks was chosen as it is unlikely that symptoms change during this interval, whereas it is long enough for the participant to forget initial responses. The exact number of days between completion of the first and second questionnaire was recorded.

Statistical analysis

IBM SPSS Statistics 20 (IBM, Armonk, New York) was used for data analysis. Descriptive statistics were applied to determine mean age, gender ratio, type of shoulder instability and days between measurements.

Measurement properties

We applied the COSMIN taxonomy to assess measurement properties of the WOSI in a systematic and comprehensive way. Mokkink et al. [16] developed the COSMIN taxonomy to clarify and standardize terminology and definitions of measurement properties to evaluate HR-QOL questionnaires. Consensus was reached by an international expert panel [19]. The COSMIN definitions which are relevant for underlying study are presented in Table 2.

Reliability

Reliability refers to the extent to which scores for patients who have not changed are the same for repeated measurements under several conditions (COSMIN definition). This domain contains the measurement properties: internal consistency, test–retest reliability and measurement error, which were all assessed for the WOSI in this study.

Internal consistency refers to whether several items that propose to measure the same general construct produce similar and correlating scores. The COSMIN expert panel defines it as "the interrelatedness among items", which is originally a definition from Cortina [20]. In the present study internal consistency was measured with Cronbach's alpha, a reliability coefficient ranging from 0 to 1, with a Cronbach's alpha of 0.7 and higher values indicating sufficient internal consistency [21]. Extremely high values of Cronbach's alpha (>0.95), however, may indicate the presence of redundant items.

Table 2 COSMIN domains, measurement properties and statistical parameters. Mokkink et al. [16, 19]

Domain	Measurement property	Statistical parameters
Reliability	Internal consistency	Cronbach's alpha (CA)
	Reliability (test–retest; inter-rater and intra-rater)	Intraclass Correlation Coefficient (ICC)
	Measurement error	Standard error of measurement (SEM)
		Smallest detectable change (SDC)
		Limits of agreement (LoA)
Interpretability[a]		Minimal important change (MIC)
		Floor and ceiling effects

The COSMIN domains validity and responsiveness are not shown in this table, because they are beyond the scope of this study

[a] Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument

Test–retest reliability concerns the degree to which repeated measurements (over time) provide similar results, also reported in literature as reproducibility, but the COSMIN steering committee prefers the term test–retest reliability. For total score and domain scores the test–retest reliability was calculated by the Intraclass Correlation Coefficient (ICC), using a two-way random effects model with an absolute agreement definition, assuming there are no systematic differences between measurements [22]. For the present study, we defined an ICC beyond 0.70 as good reliability, an ICC between 0.40 and 0.70 as moderate reliability and an ICC below 0.40 as poor reliability.

Measurement error was assessed using the standard error of measurement (SEM) and the smallest detectable change (SDC). SEM was calculated by $SD \times \sqrt{(1 - R)}$, with R = ICC and $SD = \sqrt{(\text{total variance})}$ [23]. The SEM was subsequently used to calculate the SDC by $1.96 \times \sqrt{2} \times SEM$. Changes larger than the SDC are considered to be real changes, i.e., changes beyond measurement error to indicate 95 % confidence for real change between the two assessments scores [24, 25].

To enable comparison with similar questionnaires the Reliable Change Index (RCI) was calculated, representing the SDC as a percentage of the maximum obtainable score.

Agreement also concerns the measurement error, and assesses how close the scores on the WOSI are for the 2 measurements. For this purpose, the Bland and Altman method was used by plotting the mean difference (mean D) between the two consecutive measurements against the standard deviation (SD) of this difference [26]. The 'limits of agreement' were calculated as the mean difference ± 1.96 times the SD of the differences. The Bland and Altman plot provides a visual interpretation of possible systematic variation in differences over the range of measurement, and outliers that are not revealed by regular correlation analyses.

Interpretability

Interpretability refers to the degree to which qualitative meaning can be assigned to an instrument's quantitative scores [19]. One aspect of interpretability is assessing floor and ceiling effects. We calculated floor and ceiling effects for the total WOSI score and for the domain scores of the first series of WOSI's. Maximal scores were defined as the top 90–100 % score ranges and minimal scores as 0–10 %. A percentage of >15 % of the participants scoring minimal or maximum scores was considered to be a relevant floor or ceiling effect.

Results

Measurement properties

Reliability

In Table 3 reliability parameters of the Dutch WOSI are presented in terms of internal consistency (Cronbach's alpha), test retest reliability (ICC) and measurement error (SEM, SDC and RCI).

Cronbach's alpha was 0.95 for the total WOSI score and ranged from 0.88 to 0.95 for the 4 domains, implying high internal consistency. ICC for the total WOSI score was 0.91 implicating good test–retest reliability. The domains ICC's ranged from 0.79 to 0.90, with the highest ICC for the physical symptoms domain and the lowest ICC for the domain of sports, recreation and work.

The standard error of measurement (SEM) of the total WOSI score was 130.6. As a result, the smallest detectable change (SDC) was 362.0 for the total WOSI score which is 17.3 % of the maximum obtainable score of 2100. SDC for the domains ranged from 93.7 to 128.0 for the 4 domains, which is 9.6–32.0 % from the maximum obtainable domain scores.

Figure 1 shows the Bland–Altman plots for total WOSI score and for the domain of physical symptoms. No systematic differences or any indications for consistent bias were observed between the first and second measurement. The same applies for the other three domains.

Table 3 Reliability parameters (CA and ICC) and measurement error (SDC and RCI)

	Internal consistency (CA)	Test–retest reliability (ICC with 95 % CI)	SEM	SDC	RCI (%)
WOSI total score	0.95	0.91 (0.84–0.95)	130.6	362.0	17.3
Physical symptoms	0.95	0.90 (0.83–0.94)	34.5	95.6	9.6
Sports, recreation, work	0.88	0.79 (0.66–0.87)	46.2	128.0	32.0
Lifestyle	0.89	0.81 (0.69–0.89)	41.4	114.8	28.7
Emotions	0.91	0.83 (0.73–0.90)	33.8	93.7	31.2

CA Cronbach's alpha, *ICC* intraclass correlation coefficient, *SEM* standard error of measurement, *SDC* smallest detectable change, *RCI* reliable change index

Interpretability

We assessed floor and ceiling effects as an aspect of interpretability according to the COSMIN taxonomy. For the total WOSI score <15 % of the patients obtained the maximum or minimum score range of 0–10 % (floor) and 90–100 % (ceiling), implying there were no floor and ceiling effects for the total score. A similar result was found for the 4 domain scores, with the exception of the lifestyle domain, for which 8 patients (15.3 %) obtained scores in the minimal score ranges, implying a mild floor effect (Table 4).

Discussion

The current study evaluates the measurement properties of the Dutch version of the WOSI, which we translated according to international guidelines [17]. To our knowledge the present study is the first reporting the translation and measurement properties of the Dutch WOSI. Similar to previous studies on the original WOSI and translated versions, we found good to excellent measurement properties.

Internal consistency, represented by Cronbach's alpha was excellent with 0.95 for the total WOSI score, and values between 0.88 and 0.95 for the domains. Although not reported for the original WOSI, the Swedish translation showed a similar Cronbach's alpha of 0.95 in a smaller group of 22 patients [16]. In addition, the two German versions (Hofstaetter and Drerup) and the Italian and Japanese WOSI's reported slightly lower values of Cronbach's alpha than we found (values ranging from 0.84 to 0.93) but still indicated good to excellent internal consistency [10–12, 14].

A Cronbach's alpha exceeding 0.95 might imply the presence of redundant items, but not one WOSI study reported such high values.

We found an ICC of 0.91 for the total WOSI score, implicating good test–retest reliability. For the domains, ICCs ranged from 0.79 to 0.90, which is also good (exceeding 0.70). These results are similar to those of the original WOSI and translated versions, with ICCs varying from 0.87 to 0.98 [2, 10–14]. However, there is a large variation in the number of subjects included for analysis between the studies. Only 4 studies, including the original study of Kirkley et al.. and the present study [2, 11, 12], used 50 or more patients, which is an appropriate sample size to assess reliability parameters in health status questionnaires [18].

Measurement error of the WOSI was 130.6 on a scale from 0 to 2100, expressed in SEM. This resulted in an SDC of 362 and an RCI of 17.3 %, meaning that a score difference >362 points on the WOSI indicates true improvement or impairment. Measurement error of the WOSI was recently described for the first time by Cacchio et al. [11], who cross culturally adapted the WOSI for use in Italy. They found a SEM of 71 and SDC of 196, reported as minimal detectable change (MDC) in that study. Since only one study has calculated the measurement error of the WOSI before, it is premature to draw conclusions on this part. The SDC we found appears high compared to the Italian version. However, for the Western Ontario Rotator Cuff Index (WORC), which is a questionnaire with comparable characteristics as the WOSI, similar values for SDC and RCI are reported for the Dutch and Norwegian translations [27–29].

In our study we examined a heterogeneous patient group, with mild to severe shoulder instability. The whole range of potential WOSI scores was covered (0–2100). ICC is highly dependent on the variation of the study population, where a heterogeneous group leads to higher ICCs than a homogeneous group. ICC can only be generalized to populations with similar variation [30]. In the clinical setting, patients with shoulder instability vary a lot in, e.g., frequency of dislocations, pain and functional problems. Therefore, investigating a study population with similar variation, as done in the current study, is crucial for translation of the results to the clinical setting. We found high ICC for total WOSI score (0.91) with a narrow CI (95 % CI 0.84–0.95), indicating that the Dutch WOSI is useful for group evaluation and for measuring individual change. This is confirmed by a Cronbach's alpha, exceeding 0.90, which is the recommended threshold for using HR-QOL's in the clinical setting.

The increasing interest in measurement properties of HR-QOL questionnaires has led to many publications on this topic. Despite the recent publication of the COSMIN

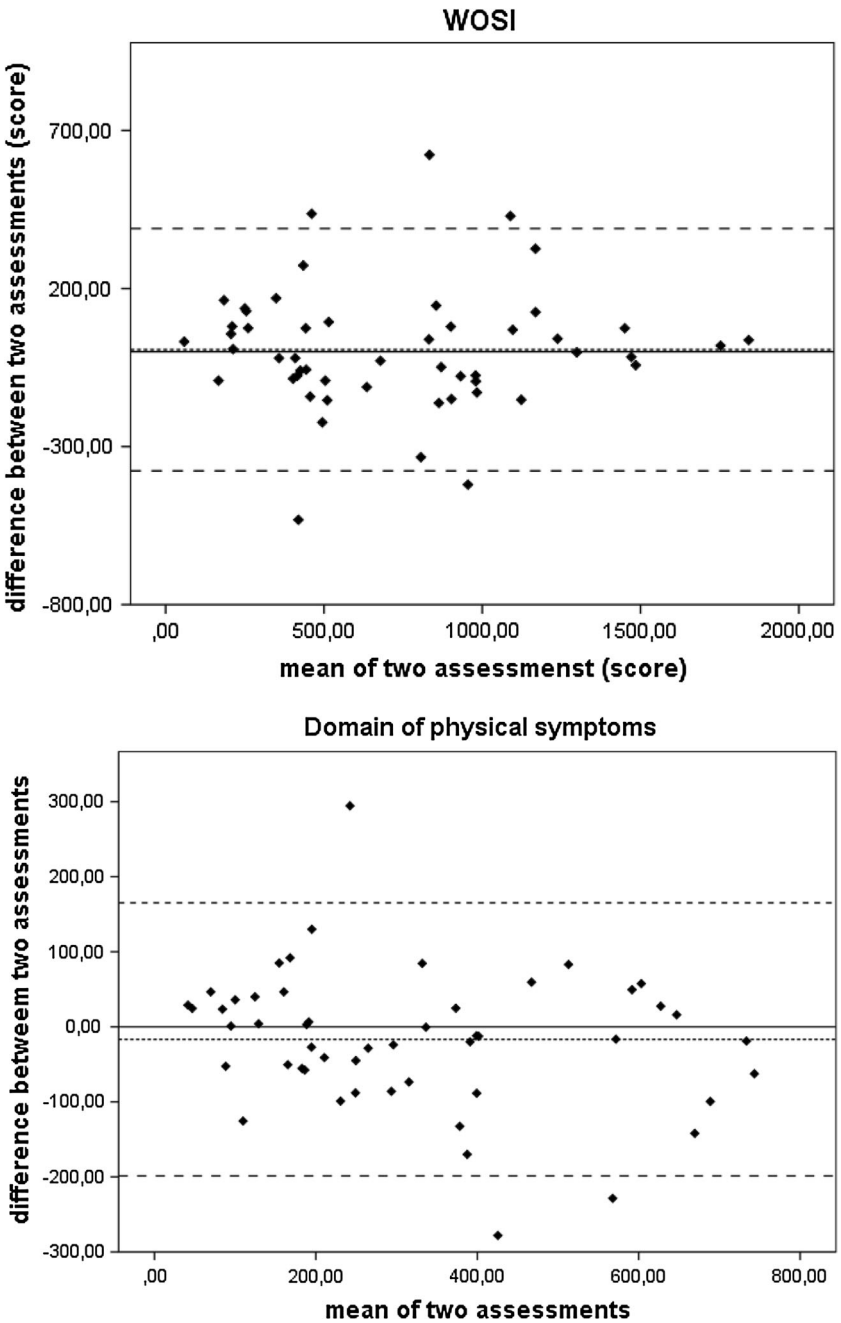

Fig. 1 Bland–Altman plots for total WOSI score and for the domain of physical symptoms. *Bold dotted line* the mean difference score. *Thin dotted lines* the limits of agreement, defined as the mean ± SD of the difference score

taxonomy to clarify and standardize definitions of measurement properties, still numerous terms and definitions are used interchangeably for the same constructs in literature [19]. We encourage researchers in the field of HR-QOL questionnaires to apply the COSMIN taxonomy in future research.

Limitations and future studies

We assessed reliability of the Dutch version of the WOSI according to the COSMIN taxonomy. We did not test the validity (comparison with other clinical scores) and responsiveness (compare clinical scores before and after an intervention) of the WOSI. However, this has been thoroughly investigated for the original and other translated WOSI questionnaires. These studies describe the WOSI as a valid questionnaire, which is highly responsive to change over time. Our results on the measurement properties of the reliability domain (internal consistency, test–retest reliability, and measurement error) are comparable with earlier studies, so it is likely that the Dutch WOSI will have similar outcomes on validity and responsiveness parameters. However additional research is required to further validate the Dutch WOSI with regard to these specific parameters.

Table 4 Floor and ceiling effects of the first measurement series

	Score ranges	Floor effect: *N* (%)	Ceiling effect: *N* (%)
WOSI total score	0–2100	N: 4 (7.7 %)	N: 0
Physical symptoms	0–1000	N: 7 (13.4 %)	N: 0
Sports, recreation, work	0–400	N: 7 (13.4 %)	N: 2 (3.8 %)
Lifestyle	0–400	Y: 8 (15.4 %)	N: 2 (3.8 %)
Emotions	0–300	N: 6 (11.5 %)	N: 5 (9.6 %)

N (%) number and percentage of patients obtaining minimum (0–10 %) or maximum (90–100 %) score range

N no, *Y* yes

A factor analysis is commonly done before the Cronbach's alpha is calculated. However the sample size of 52 patients was too small to perform a significant factor analysis.

The mean interval between the measurements was 25 days, which is longer than the 2 weeks described in the protocol. The median interval was 20 days. Because we included patients who achieved a stable situation after trauma or surgery, a slightly longer interval is preferable over a shorter interval, because a shorter interval bears the risk of not forgetting the initial response. We found high ICCs despite the longer time interval, indicating that our study population actually remains stable during the measurement period.

Conclusion

The results of the present study suggest the Dutch version of the WOSI is a reliable tool for clinical assessment and scientific evaluation. It shows high values for Cronbach's alpha and ICC implying excellent internal consistency and good test–retest reliability.

Acknowledgments We thank Sharon Griffin for her permission to translate the original WOSI, Arthur de Gast for his participation in the translation process and Caroline Terwee for her assistance in interpreting COSMIN correctly.

Conflict of interest The authors, their immediate family, and any research foundation with which they are affiliated did not receive any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. van der Heijden GJ. Shoulder disorders: a state-of-the-art review. Baillieres Best Pract Res Clin Rheumatol. 1999;13:287–309.
2. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med. 1998;26:764–72.
3. Kuhn JE. A new classification system for shoulder instability. Br J Sports Med. 2010;44:341–6.
4. Hayes K, Callanan M, Walton J, Paxinos A, Murrell GA. Shoulder instability: management and rehabilitation. J Orthop Sports Phys Ther. 2002;32:497–509.
5. Simmen BR, Angst F, Schwyzer HK, Herren DB, Pap G, Aeschlimann A, Goldhahn J. A concept for comprehensively measuring health, function and quality of life following orthopaedic interventions of the upper extremity. Arch Orthop Trauma Surg. 2009;129:113–8.
6. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. 1989;27(3 Suppl):S217–32.
7. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis. 2004;63:335–41.
8. Wright RW, Baumgarten KM. Shoulder outcomes measures. J Am Acad Orthop Surg. 2010;18:436–44.
9. Rouleau DM, Faber K, MacDermid JC. Systematic review of patient-administered shoulder functional scores on instability. J Should Elb Surg. 2010;19:1121–8.
10. Hofstaetter JG, Hanslik-Schnabel B, Hofstaetter SG, Wurnig C, Huber W. Cross-cultural adaptation and validation of the German version of the Western Ontario Shoulder Instability Index. Arch Orthop Trauma Surg. 2010;130:787–96.
11. Cacchio A, Paoloni M, Griffin SH, Rosa F, Properzi G, Padua L, Padua R, Carnelli F, Calvisi V, Santilli V. Cross-cultural adaptation and measurement properties of an Italian version of the Western Ontario Shoulder Instability Index (WOSI). J Orthop Sports Phys Ther. 2012;42:559–67.
12. Hatta T, Shinozaki N, Omi R, Sano H, Yamamoto N, Ando A, Sugaya H, Aizawa T, Kuriyami S, Itoy E. Reliability and validity of the Western Ontario Shoulder Instability Index (WOSI) in the Japanese population. J Orthop Sci. 2011;16:732–6.
13. Salomonsson B, Ahlstrom S, Dalen N, Lillkrona U. The Western Ontario Shoulder Instability Index (WOSI): validity, reliability, and responsiveness retested with a Swedish translation. Acta Orthop. 2009;80:233–8.
14. Drerup S, Angst F, Griffin S, Flury MP, Simmen BR, Goldhahn J. Western Ontario shoulder instability index (WOSI): translation and cross-cultural adaptation for use by German speakers. Orthopade. 2010;39:711–8.
15. van der Linde JA, van Kampen DA, Terwee CB, Dijksman LM, Kleinjan G, Willems WJ. Long-term results after arthroscopic shoulder stabilization using suture anchors: an 8- to 10-year follow-up. Am J Sports Med. 2011;39:2396–403.
16. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter L, de Vet H. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19:539–49.
17. Acquadro C, Jambon B, Ellis D, Marquis P. Language and translation issues. Philadelphia: Lippincot-Raven; 1996.
18. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter L, de Vet H. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter L, de Vet H. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–45.

20. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol. 1993;78:98–104.
21. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
22. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil. 1998;12:187–99.
23. Weir JP. Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19:231–40.
24. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res. 2001; 10:571–8.
25. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10.
27. Wiertsema SH, Rietberg MB, Hekman KM, Schothorst M, Steultjens MP, Dekker J. Reproducibility of the Dutch version of the Western Ontario Rotator Cuff Index. J Should Elb Surg. 2013;22:165–70.
28. de Witte PB, Henseler JF, Nagels J, Vliet Vlieland TP, Nelissen RG. The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. Am J Sports Med. 2012;40:1611–9.
29. Ekeberg OM, Bautz-Holter E, Tveita EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. BMC Musculoskelet Disord. 2008;9:68.
30. Streiner DL, Norman GR. Health measurement scales. Oxford: Medical Publication; 2008.